

(19) 日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2002-189499
(P2002-189499A)

(43) 公開日 平成14年7月5日 (2002.7.5)

(51) Int.Cl.⁷G 10 L 19/02
19/00
H 03 M 7/30

識別記号

F I

H 03 M 7/30
G 10 L 7/04
9/18テマコード (参考)
A 5D045
G 5J064
H
M

審査請求 有 請求項の数4 O.L (全5頁)

(21) 出願番号

特願2000-387351(P2000-387351)

(22) 出願日

平成12年12月20日 (2000.12.20)

(71) 出願人 000004075

ヤマハ株式会社

静岡県浜松市中沢町10番1号

(72) 発明者 鈴木 典雄

静岡県浜松市中沢町10番1号 ヤマハ株式
会社内

(74) 代理人 100064908

弁理士 志賀 正武 (外1名)

Fターム(参考) 5D045 DA06

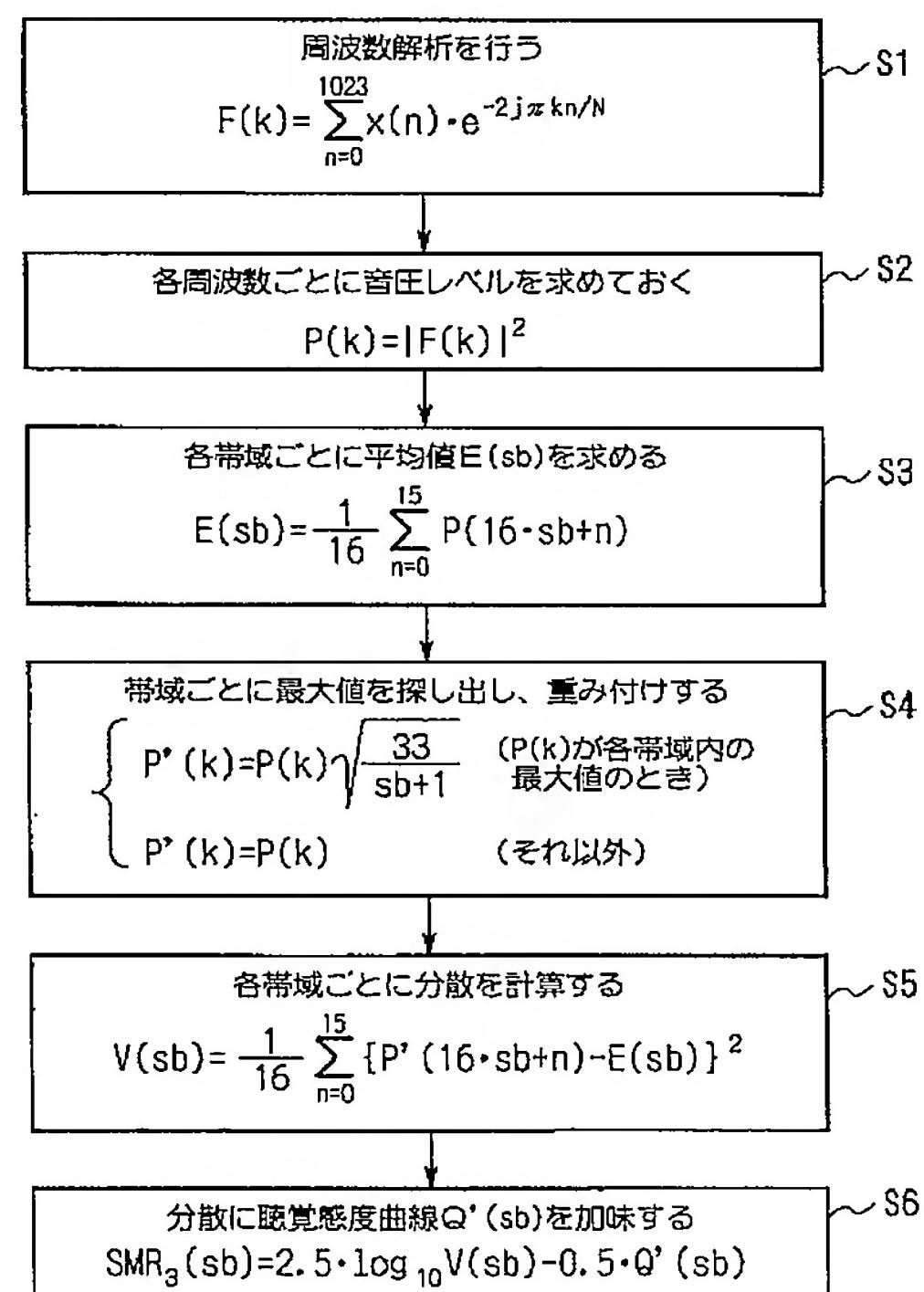
5J064 AA03 BA16 BB12 BC08 BC09
BC11 BC16 BC21 BC22 BC25
BC28 BC29

(54) 【発明の名称】 デジタルオーディオ信号圧縮方法および圧縮装置

(57) 【要約】

【課題】 計算時間を従来より大幅に短縮することができるデジタルオーディオ信号圧縮方法を提供する。

【解決手段】 この圧縮方法は、まず、各サブバンド毎に周波数成分を求め (ステップS1) 、次に、その周波数成分を音圧レベルに変換し (ステップS2) 、次に、各サブバンド毎に音圧レベルの平均値を計算し (ステップS3) 、次に、各サブバンド毎に最大音圧レベルの重み付けを行い (ステップS4) 、次に、各サブバンド毎に音圧レベルの分散を計算し (ステップS5) 、次に、聴覚感度データによる補正を行ってSMRを求める (ステップS6) 。求めたSMRに基づいてビット割り当てを行い、ビット割り当てに従って各サブバンドの信号を量子化して出力する。



【特許請求の範囲】

【請求項1】 入力ディジタルオーディオ信号に対して複数のサブバンドに周波数分割すると共に、心理聴覚分析処理を行い、該心理聴覚分析処理の結果に従って各サブバンドのビット割り当てを行い、ビット割り当てに従って各サブバンドの信号を量子化して出力するディジタルオーディオ信号圧縮方法において、

前記心理聴覚分析処理は、入力ディジタルオーディオ信号を周波数解析して周波数成分に変換し、次に、各サブバンド毎の前記周波数成分の分散を計算する処理であることを特徴とするディジタルオーディオ信号圧縮方法。

【請求項2】 前記心理聴覚分析処理は、入力ディジタルオーディオ信号を周波数解析して周波数成分に変換した後、各サブバンド毎の周波数成分の最大値を検出し、検出した最大値にそのサブバンドの周波数帯域に応じた重み付けを行い、重み付けを行った最大値を用いて前記分散を計算することを特徴とする請求項1に記載のディジタルオーディオ信号圧縮方法。

【請求項3】 前記心理聴覚分析処理は、前記分散を計算した後、該計算によって得られた値に、求められている聴覚感度に対応するデータを演算することを特徴とする請求項1または請求項2に記載のディジタルオーディオ信号圧縮方法。

【請求項4】 入力ディジタルオーディオ信号に対して複数のサブバンドに周波数分割すると共に、心理聴覚分析処理を行う心理聴覚分析手段と、該心理聴覚分析処理の結果に従って各サブバンドのビット割り当てを行うビット割り当て手段と、該ビット割り当てに従って各サブバンドの信号を量子化して出力する量子化手段とを具備するディジタルオーディオ信号圧縮装置において、前記心理聴覚分析手段は、入力ディジタルオーディオ信号を周波数解析して周波数成分に変換する第1の手段と、各サブバンド毎の前記周波数成分の分散を計算する第2の手段とを具備することを特徴とするディジタルオーディオ信号圧縮装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 この発明は、MPEG/ Audio規格、ATRAC規格、ドルビーディジタル規格等によるオーディオ信号圧縮において用いられるディジタルオーディオ信号圧縮方法および圧縮装置に関する。

【0002】

【従来の技術】 図3は、MPEG(Moving Picture Coding Experts Group)/Audio規格によるディジタルオーディオ信号圧縮回路の構成を示す回路図である。この図において、入力ディジタルオーディオ信号D_aは所定のサンプル数毎にブロック化(フレームと言う)され、2つのパ*

*スに分かれて処理される。まず、一方のパスにおけるファイルタバンク1は、入力信号を等しい帯域幅を持つ32バンドのサブバンド信号に分割する。この場合、各々のサブバンド信号は1/32のサンプリング周波数にダウンサンプルされる。スケールファクタ抽出・正規化回路2は、1フレームにおける各々のサブバンド信号に対し、最大絶対値を持つサンプルを検出する。その値を対数に変換し量子化したものとスケールファクタと呼ぶ。そして、このスケールファクタによって各サブバンドサンプルを除算し、それらの値を±1の範囲内に正規化する。

【0003】一方、心理聴覚分析部3は、FFT(高速フーリエ変換)による周波数スペクトルの計算を行い、それによって得られた各周波数データに基づき各サブバンド毎のマスキングしきい値を計算して出力する。ビット割当部4は心理聴覚分析部3の出力と、ビットレートで決まる1フレームで使用可能なビット数の制限の下で反復ループ処理により各サブバンド毎の量子化ビット数を決定する。量子化部5は、各サブバンド毎に設定された量子化ビット数でスケールファクタ抽出・正規化回路2から出力されるサブバンド信号を量子化する。ビットストリーム生成部6は、量子化されたサブバンドサンプル、各サブバンドに対するビット割当情報およびスケールファクタをマルチプレックスし、それにヘッダを付けてビットストリームを作成し出力する。

【0004】次に、従来の心理聴覚分析部3における処理手順の一例を説明する。なお、以下に説明する手順は、ISO/IEC 11172-3による心理聴覚モデルにおけるMODEL1の手順である。

- 30 (1) FFTによって周波数特性を求め、512の周波数データを得る。
- (2) 32のサブバンドの各音圧レベルを求める。
- (3) 絶対しきい値を決定する。
- (4) 音として聞こえる周波数(マスク)を選ぶ。
- (5) マスクを減らす。
- (6) 各々のマスクしきい値を計算する。
- (7) 大域的マスクしきい値を計算する。
- (8) 各サブバンドの最小マスクしきい値を決定する。
- (9) 各サブバンドの信号対マスク比(SMR)を計算する。そして、上記SMRがビット割当情報としてビット割当部4へ出力される。

【0005】

【発明が解決しようとする課題】 ところで、上述した従来の心理聴覚分析部3における処理は、計算に時間がかかる欠点があった。特に、上記(6)、(7)の計算に時間がかかっていた。例えば、上記(7)の計算は、

【数1】

$$LTg(i) = 10 \log_{10} \left(10^{LTg(i)/10} + \sum_{j=1}^m 10^{LTlm(z(j), z(i))/10} + \sum_{j=1}^n 10^{LTnm(z(j), z(i))/10} \right)$$

なる式の計算であり、 10^g の計算を除外したとしても、 i は 1 ~ 約 130 であり、 m 、 n は音の数とノイズの数で 10 ~ 20 程度あり、したがって、 $L T_{tm}$ と $L T_{nm}$ をそれぞれ 1000 回以上計算しなくてはならない。 $L T_{tm}$ と $L T_{nm}$ は 3 項の和で表現されるが、その内 2 項は 1 次関数となっており、計算に時間がかかる。

【0006】この発明は、このような事情を考慮してなされたもので、その目的は、計算時間を従来より大幅に短縮することができるディジタルオーディオ信号圧縮方法および圧縮装置を提供することにある。

【0007】

【課題を解決するための手段】この発明は上記の課題を解決すべくなされたもので、請求項 1 に記載の発明は、入力ディジタルオーディオ信号に対して複数のサブバンドに周波数分割すると共に、心理聴覚分析処理を行い、該心理聴覚分析処理の結果に従って各サブバンドのビット割り当てを行い、ビット割り当てに従って各サブバンドの信号を量子化して出力するディジタルオーディオ信号圧縮方法において、前記心理聴覚分析処理は、入力ディジタルオーディオ信号を周波数解析して周波数成分に変換し、次に、各サブバンド毎の前記周波数成分の分散を計算する処理であることを特徴とするディジタルオーディオ信号圧縮方法である。

【0008】また、請求項 2 に記載の発明は、請求項 1 に記載のディジタルオーディオ信号圧縮方法において、前記心理聴覚分析処理は、入力ディジタルオーディオ信号を周波数解析して周波数成分に変換した後、各サブバンド毎の周波数成分の最大値を検出し、検出した最大値にそのサブバンドの周波数帯域に応じた重み付けを行い、重み付けを行った最大値を用いて前記分散を計算することを特徴とする。また、請求項 3 に記載の発明は、請求項 1 または請求項 2 に記載のディジタルオーディオ信号圧縮方法において、前記心理聴覚分析処理は、前記分散を計算した後、該計算によって得られた値に、求められている聴覚感度に対応するデータを演算することを特徴とする。

【0009】また、請求項 4 に記載の発明は、入力ディジタルオーディオ信号に対して複数のサブバンドに周波数分割すると共に、心理聴覚分析処理を行う心理聴覚分析手段と、該心理聴覚分析処理の結果に従って各サブバンドのビット割り当てを行うビット割り当て手段と、該ビット割り当てに従って各サブバンドの信号を量子化して出力する量子化手段とを具備するディジタルオーディオ信号圧縮装置において、前記心理聴覚分析手段は、入力ディジタルオーディオ信号を周波数解析して周波数成分に変換する第 1 の手段と、各サブバンド毎の前記周波数成分の分散を計算する第 2 の手段とを具備することを特徴とするディジタルオーディオ信号圧縮装置である。

【0010】

【発明の実施の形態】以下、図面を参照し、この発明の

一実施の形態について説明する。この実施形態による圧縮方法を適用した圧縮回路は図 3 と同じであり、以下に説明する処理手順は、図 3 の心理聴覚分析部 3 において用いられる。図 1 は同実施の形態による圧縮方法を説明するためのフローチャートである。以下、各ステップ S 1 ~ S 6 を順次説明する。なお、以下の説明においては、サブバンドの数を 32、各サブバンド内の周波数データの数を 16 とする。

【0011】○ステップ S 1 (周波数解析)

10 入力されるディジタルオーディオデータ（実数）に対し、FFT 処理（ステップ S 1 に記載される式の演算）を行い、MPEG/Audio Layer2(MP2)の場合、周波数データ $F(k)$ として、1024 データを得る。この周波数データ $F(k)$ は複素数データであり、実数入力の FFT の対象性により、意味があるのは 512 データである。なお、図の式において j は虚数単位である。

【0012】○ステップ S 2 (音圧レベル測定)

各周波数データ $F(k)$ の実部と虚部の二乗の和、すなわち $F(k)$ の絶対値の二乗 $P(k)$ を求める。この値 $P(k)$ が音圧レベルに相当する。

○ステップ S 3 (平均値の測定)

20 上述した 512 の音圧データを 16 データ毎に 32 のサブバンド（帯域）に分ける。そして、各サブバンド毎に、音圧レベル $P(k)$ の平均値 $E(s_b)$ を求める。なお、 s_b はサブバンド番号であり、低音側から 0 ~ 31 が割り当てられている。例えば、 $s_b = 3$ のサブバンドには、 $P(48) \sim P(63)$ の音圧レベルが含まれる。

【0013】○ステップ S 4 (最大値の重み付け)

30 各サブバンド毎に音圧レベル $P(k)$ の最大値 $P'(k)$ を検出し、検出した音圧レベル $P'(k)$ に $\sqrt{(33 / (s_b + 1))}$ {sqrt : 二乗根} なる値を乗算することによって、重み付けした音圧レベル $P'(k)$ を得る。

○ステップ S 5 (分散の計算)

各サブバンド毎に、上述したステップ S 2 ~ S 4 の演算結果を用いて、音圧レベル $P(k)$ の分散 $V(s_b)$ を計算する。ここで、各帯域の最大音圧レベル $P(k)$ については、上記重み付けした音圧レベル $P'(k)$ を用いる。

【0014】○ステップ S 6 (SMR の計算)

各サブバンドの SMR を計算する。すなわち、各サブバンドについて、分散 $V(s_b)$ の対数をとり、2.5 倍し、その値から聴覚感度データ $Q'(s_b)$ の 0.5 倍を減算する。ここで、聴覚感度データ $Q'(s_b)$ とは、人間の耳の聴覚感度曲線に対応したデータであり、予めメモリ内に記憶させておく。図 2 に感度聴覚データの一例を示す。この図において「 F_s 」は、アナログオーディオ信号をディジタルオーディオ信号に変換する際のサンプリング周波数である。

【0015】以上がこの発明の実施形態による処理手順である。上述したことから明らかなように、この実施形態による心理聴覚分析部3のSMR計算方法は、基本的には、各サブバンド毎に音圧レベル $P(k)$ の分散を求め、求めた分散値をSMRとして使用するという方法である。ここで、分散がSMRとして使用できる理由は次の通りである。すなわち、各サブバンドの16の音圧レベルの平均値が同じであったとしても、バラツキが大きければ分散は大きな値となり、逆に、バラツキが小さければ分散は小さな値となる。全ての音圧レベルが等しければ分散は0となる。一方、1つのサブバンドにおける周波数成分にバラツキが少ない場合、それぞれの周波数成分の位相がまちまちであれば、人間の耳には雑音に聞こえる。これに対し、サブバンド内にピークがあれば、その音が知覚される。すなわち、周波数成分（振幅または音圧レベル）の分散を計算した時に、分散が小さい波形より大きい波形の方がより重要な波形であり、したがって、ビット数も多くすることが必要となり、一方、分散が小さい波形はビット数を少なくすることができます。

【0016】ところで、SMRと分散には以下の性質がある。

- (1)データ量割り当て1(ビット)がSMRの約6(dB)に相当する。
- (2)分散は音圧レベルの二乗の次元を持っている。
- (3)音圧レベルは振幅の二乗の次元を持っている。
- (4)データ量は振幅が2倍になった時に1(ビット)増えやすのが妥当である。すなわち、分散を振幅に直すために4乗根をとり、さらに、常用対数をとって20倍(dBへの変換)してSMRとするのが妥当である。実際には、対数の性質から、分散の常用対数をとって5倍すればよい。つまり、SMRは、基本的には、 $SMR(s_b) = 5 \times \log_{10} V(s_b)$ なる式によって求められる。

【0017】しかし、人間の耳には各種聴覚特性があるため、その特性を利用した方がよりよい結果が得られる。まず、人間の耳は低域ほど周波数分解能がよく、高域になるに従い分解能が悪くなる性質がある。この性質を考慮し、次の補正方法が考えられる。分散を計算する時、各サブバンドにおける最大の音圧レベル $P(k)$ のみ周波数に反比例させて大きくする。すなわち、各サブバンドにおける最大の音圧レベルを $P'(k)$ とすると、この $P'(k)$ に、

$$P'(k) = P(k) \times (32/s_b)$$

なる補正を加え、この補正後の音圧レベル $P'(k)$ を使用して分散を計算する。

【0018】実験的には、直接反比例させると高域特性が著しく劣化するため、二乗根に反比例させた方が良い結果が得られる。また、 s_b は0から始まるため計算の都合上「1」を加算する。結局、次式によって音圧レベルを補正することが好ましい。

$P'(k) = P(k) \times \sqrt{33/(s_b + 1)}$

前述したステップS4の重み付けはこの補正である。なお、平均値の計算（ステップS3）においては、補正前の音圧レベル $P(k)$ を使用する。

【0019】次に、人間の耳にはいわゆる聴覚曲線で表される周波数特性がある。各サブバンドの中央における聴覚感度を $Q(s_b)$ とすると、この値は音圧レベル(dB)単位で表されており、値が小さい（負にもなる）ほど、耳の感度が良いことを示している。そこで、この聴覚感度 $Q(s_b)$ を加味した次式によってSMRを計算する方が好ましい。

$$SMR_2(s_b) = 5 \times \log_{10} V(s_b) - Q(s_b)$$

【0020】しかし、実際には、例えばMP2でサンプリング周波数48KHzの時を考えてみると、 $s_b = 0$ に相当する周波数は0～750Hzであり、この範囲には多くの音の基本波が存在するため、感度を落とすと音質が低下し、ゴソゴソという感じのノイズが増加する。したがって、耳の感度がある程度良くなる2KHz程度までは補正をしない方がよい結果が得られる。そのように修正した値を $Q'(s_b)$ とする。図2はこの $Q'(s_b)$ の値を示す。また、実験的には、単純に和をとるより、0.5:0.5の割合で和をとった方が良い結果が得られる。すなわち、次式によってSMRを計算した方がより好ましい。

$$SMR_3(s_b) = 2.5 \times \log_{10} V(s_b) - 0.5 Q'(s_b)$$

上述したステップS6はこの計算処理を示している。

【0021】以上詳述したように、上記実施形態によれば、各サブバンド毎に分散を計算し、この計算結果からSMRを求めているので、従来の演算（[数1]参照）よりはるかに簡単な演算によってSMRを求めることができる。実験では、従来の方法の約1/3の時間でSMRを求めることができた。なお、上記実施形態による方法は、従来の心理聴覚モデルに比較し、ビットレートが下がると、高域の特性が早く落ちる傾向がある。この結果、従来の心理聴覚モデルでは目立つ高域の「ピロピロ」という感じの量子化ノイズがこの実施形態による方法では目立たず、代わりに、低域の「ゴソゴソ」という量子化ノイズが目立つようになる。

【0022】

【発明の効果】以上説明したように、この発明によれば、各サブバンド毎に周波数成分を求め、求めた周波数成分の分散を計算し、この計算結果に基づいてビット割り当てを行うので、ビット割り当てのための計算時間を従来より大幅に短縮することができる利点が得られる。

【図面の簡単な説明】

【図1】この発明による方法の一実施形態の処理手順を示すフローチャートである。

【図2】同実施形態において用いられる聴覚感度デー

タを示す図である。

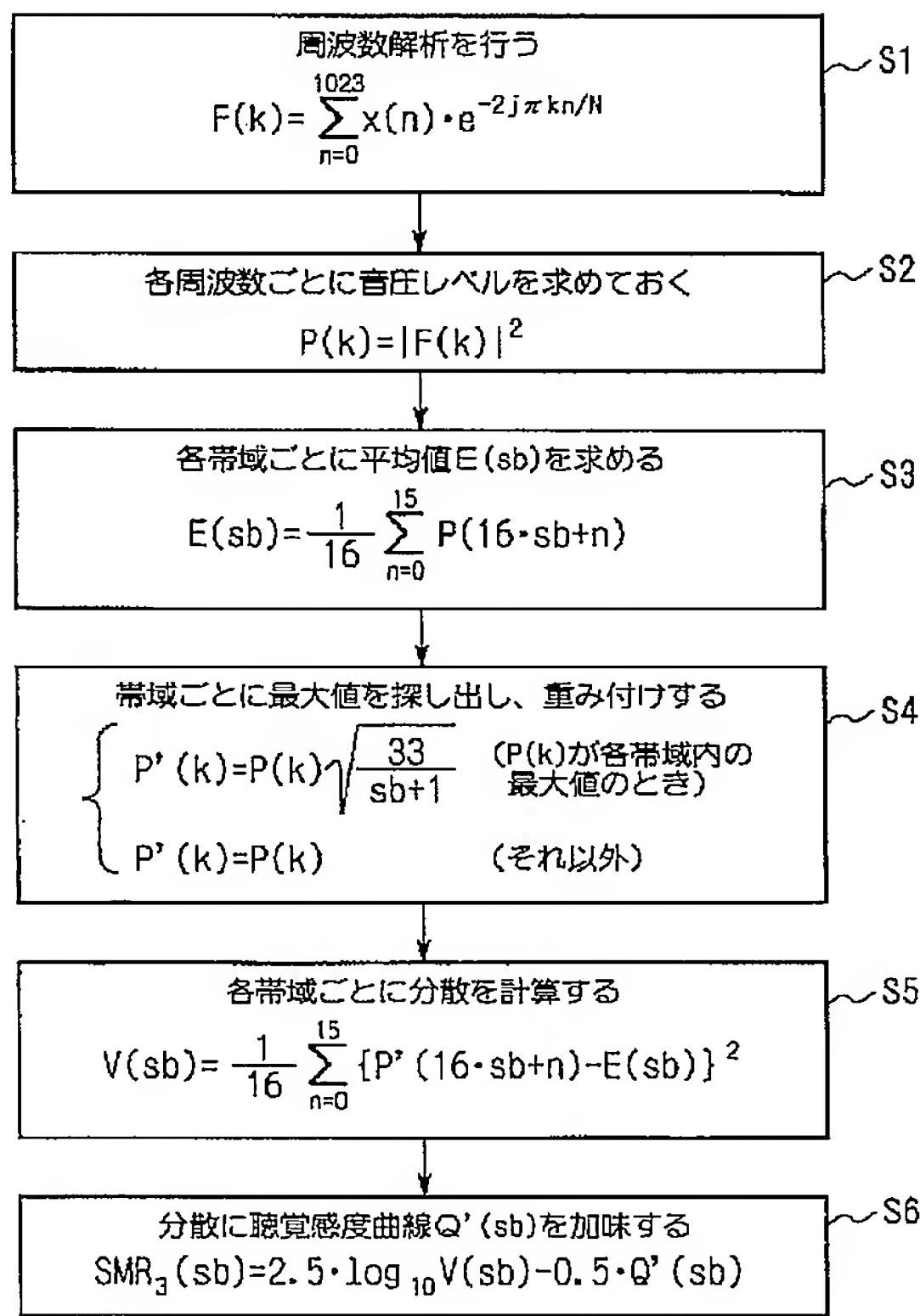
【図3】 ディジタルオーディオ信号圧縮回路の構成例
を示すブロック図である。

*

* 【符号の説明】

3…心理聴覚分析部、4…ビット割当部、5…量子化回路。

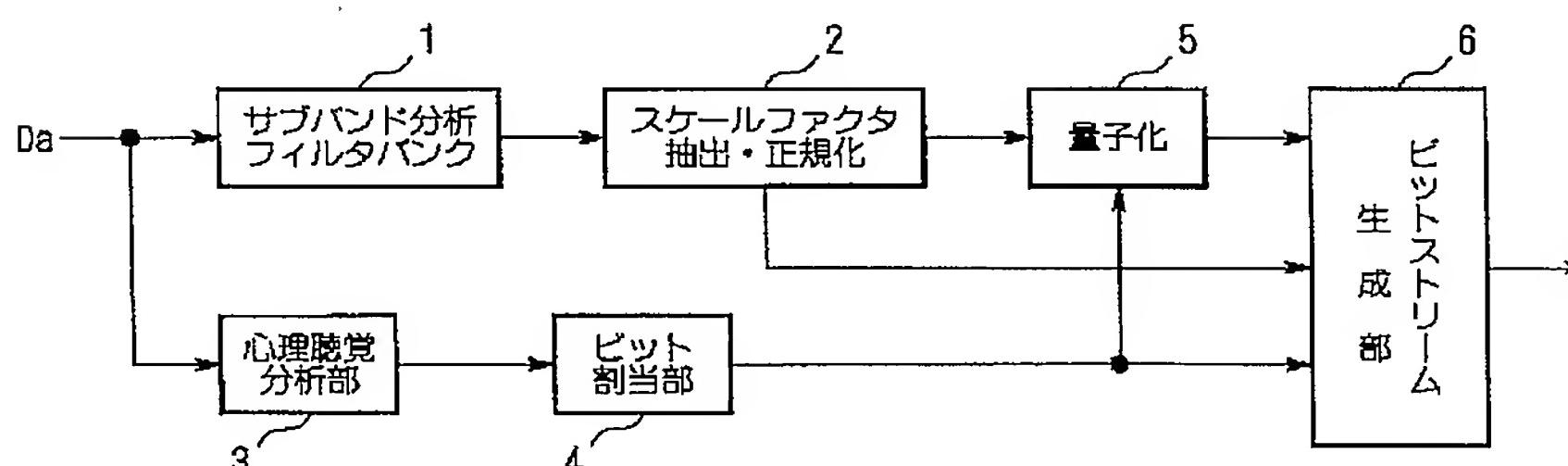
【図1】



【図2】

| 1. Fs=32KHz | | | 2. Fs=44.1KHz | | | 3. Fs=48KHz | | |
|-------------|--------|------------|---------------|--------|------------|-------------|--------|------------|
| sb | f[Hz] | Q'(sb)[dB] | sb | f[Hz] | Q'(sb)[dB] | sb | f[Hz] | Q'(sb)[dB] |
| 0 | 0 | 0.00 | 0 | 0 | 0.00 | 0 | 0 | 0.00 |
| 1 | 500 | 0.00 | 1 | 689 | 0.00 | 1 | 750 | 0.00 |
| 2 | 1,000 | 0.00 | 2 | 1,378 | 0.00 | 2 | 1,500 | 0.00 |
| 3 | 1,500 | 0.00 | 3 | 2,067 | -0.56 | 3 | 2,250 | -1.43 |
| 4 | 2,000 | -0.25 | 4 | 2,756 | -3.77 | 4 | 3,000 | -4.57 |
| 5 | 2,500 | -2.64 | 5 | 3,445 | -4.92 | 5 | 3,750 | -4.29 |
| 6 | 3,000 | -4.57 | 6 | 4,134 | -2.82 | 6 | 4,500 | -1.24 |
| 7 | 3,500 | -4.86 | 7 | 4,823 | -0.04 | 7 | 5,250 | 1.06 |
| 8 | 4,000 | -3.39 | 8 | 5,512 | 1.51 | 8 | 6,000 | 2.08 |
| 9 | 4,500 | -1.24 | 9 | 6,201 | 2.28 | 9 | 6,750 | 2.86 |
| 10 | 5,000 | 0.48 | 10 | 6,890 | 3.03 | 10 | 7,500 | 3.89 |
| 11 | 5,500 | 1.49 | 11 | 7,579 | 4.02 | 11 | 8,250 | 5.31 |
| 12 | 6,000 | 2.08 | 12 | 8,268 | 5.35 | 12 | 9,000 | 7.19 |
| 13 | 6,500 | 2.59 | 13 | 8,957 | 7.07 | 13 | 9,750 | 9.63 |
| 14 | 7,000 | 3.17 | 14 | 9,646 | 9.25 | 14 | 10,500 | 12.71 |
| 15 | 7,500 | 3.89 | 15 | 10,335 | 11.97 | 15 | 11,250 | 16.54 |
| 16 | 8,000 | 4.79 | 16 | 11,025 | 15.31 | 16 | 12,000 | 21.23 |
| 17 | 8,500 | 5.88 | 17 | 11,714 | 19.34 | 17 | 12,750 | 26.90 |
| 18 | 9,000 | 7.19 | 18 | 12,403 | 24.15 | 18 | 13,500 | 33.67 |
| 19 | 9,500 | 8.75 | 19 | 13,092 | 29.84 | 19 | 14,250 | 41.67 |
| 20 | 10,000 | 10.58 | 20 | 13,781 | 36.52 | 20 | 15,000 | 51.04 |
| 21 | 10,500 | 12.71 | 21 | 14,470 | 44.27 | 21 | 15,750 | 61.94 |
| 22 | 11,000 | 15.18 | 22 | 15,189 | 53.22 | 22 | 16,500 | 68.00 |
| 23 | 11,500 | 18.01 | 23 | 15,848 | 63.49 | 23 | 17,250 | 68.00 |
| 24 | 12,000 | 21.23 | 24 | 16,537 | 68.00 | 24 | 18,000 | 68.00 |
| 25 | 12,500 | 24.90 | 25 | 17,226 | 68.00 | 25 | 18,750 | 68.00 |
| 26 | 13,000 | 29.03 | 26 | 17,915 | 68.00 | 26 | 19,500 | 68.00 |
| 27 | 13,500 | 33.67 | 27 | 18,604 | 68.00 | 27 | 20,250 | 68.00 |
| 28 | 14,000 | 38.86 | 28 | 19,293 | 68.00 | 28 | 21,000 | 68.00 |
| 29 | 14,500 | 44.63 | 29 | 19,982 | 68.00 | 29 | 21,750 | 68.00 |
| 30 | 15,000 | 51.04 | 30 | 20,671 | 68.00 | 30 | 22,500 | 68.00 |
| 31 | 15,500 | 68.00 | 31 | 21,360 | 68.00 | 31 | 23,250 | 68.00 |

【図3】



(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: 2002189499 A

(43) Date of publication of application: 05.07.02

(51) Int. Cl

G10L 19/02
G10L 19/00
H03M 7/30

(21) Application number: 2000387351

(71) Applicant: YAMAHA CORP

(22) Date of filing: 20.12.00

(72) Inventor: SUZUKI NORIO

(54) METHOD AND DEVICE FOR COMPRESSING
DIGITAL AUDIO SIGNAL

(57) Abstract:

PROBLEM TO BE SOLVED: To provide a method for compressing digital audio signals which can make the computation time drastically shorter than before.

SOLUTION: In this compression method, the frequency component is determined first for each subband (Step S1). Then the frequency component is converted into a sound pressure level (Step S2), the average of the sound pressure level is calculated for each subband (Step S3), weighting of the maximum sound pressure level is performed for each subband (Step S4), the dispersion of the sound pressure level is calculated for each subband (Step S5), and correction by auditory sense data is performed and the SMR (signal-to-mask ratio) is determined (Step S6). Bit assignment is performed on the basis of the determined SMR, and the signal of each subband is quantized and outputted according to the bit assignment.

COPYRIGHT: (C)2002,JPO

